

**APOSTILA**

**PRIMEIRA APOSTILA AO TERMO DE EXECUÇÃO DESCENTRALIZADA - TED/UNB-001/2021 - (PROJETO SABIÁ - PROCESSAMENTO DE LINGUAGEM NATURAL APLICADO AO SISTEMA BEM-TEVI PARA ANÁLISES DE PROCESSOS JURÍDICOS DO TST). (PROCESSOS SEI/TST 600309/2021-00 E SEI/UNB 7424869)**

Por esta apostila, registra-se as alterações do Plano de Trabalho do Termo de Execução Descentralizada em epígrafe, celebrado entre o Tribunal Superior do Trabalho - TST, como Unidade Descentralizadora, e a UNIVERSIDADE DE BRASÍLIA UnB, como Unidade Descentralizada, com fundamento no § 2º, art. 15, do Decreto Lei nº 10.426, de 16 de julho de 2020, passando o Plano de Trabalho anexo ao TED a vigorar conforme segue:

## 1. Tipo de Projeto

☒ Pesquisa    ☐ Curso de Pós- Graduação    ☐ Atividade de Extensão    ☐ Ensino de Graduação

## 2. Tipo de Instrumento Processual

☐ Contrato    ☐ Convênio    ☒ Termo de Execução Descentralizada    ☐ Acordo de Cooperação    ☐ N/A

## 3. Dados Cadastrais do Proponente

Órgão/Entidade UNIVERSIDADE DE BRASÍLIA			C.N.P.J 00.038.174/0001-43	
Endereço CAMPUS UNIVERSITÁRIO DARCY RIBEIRO – PRÉDIO DA REITORIA - ASA NORTE				
Cidade BRASÍLIA	UF DF	CEP 70910-900	Telefone (61) 3107-0246	UG / Cód. Gestão 154040 / 15257
Banco Banco do Brasil - 001	Agência 1607-1		Conta Corrente 170.500-8	Praça de Pagamento Brasília
Nome do Representante Legal Márcia Abrahão Moura				CPF 334.590.531-00
CI / Órgão Exp. 960490 SSP/DF		Cargo Professora	CI / Órgão Exp. 960490 SSP/DF	Matrícula UnB 145378
Nome do Coordenador Projeto Nilton Correia da Silva				CPF 664.762.751-00
Unidade/Departamento Faculdade de Engenharias Gama/FGA				Matrícula UnB 1062816
Endereço Eletrônico (e-mail) niltoncs@unb.br / nilton.correia.silva@gmail.com			Telefone fixo (61) 3107-8901	Telefone celular (62) 99249-1953
Assinatura				
Nome do Gestor do Projeto Fabrício Ataides Braz				CPF 827.666.571-34
Unidade/Departamento Faculdade de Engenharias/FGA				Matrícula UnB 1062867
Endereço Eletrônico (e-mail) fabraz@unb.br / fabricio.braz@gmail.com			Telefone fixo (61) 3107-8901	Telefone celular (061) 98405-5045
Assinatura				

Nome do Gestor Substituto do Projeto Debora Bonat		CPF 877.397.399-87
Unidade/Departamento Faculdade de Direito		Matrícula UnB 1114441
Endereço Eletrônico (e-mail) deborabonat@unb.br	Telefone fixo (61) 3107-0714	Telefone celular (61) 98405-5045
Assinatura		

#### 4. Dados Cadastrais do Concedente

Tipo (X) Público ( ) Privado		2 – Nome / Razão Social: Tribunal Superior do Trabalho - TST		3 - CNPJ 00.509.968/0001-48	
Endereço sede (Av., Rua, Nº, Bairro) Setor de Administração Federal Sul (SAFS) Quadra 8 - Lote 1					
Cidade Brasília		UF DF	CEP 70.070 - 943	{DDD} Telefone (61) 3043-4300	{DDD} Fax
Nome do representante legal Leílio Bentes Corrêa					CPF 334.824.381-53
CI / Órgão Exp. / Emissão 652.155Órgão expedidor: SSP-DF data de expedição: 1/2/1985			Cargo Ministro Presidente do Tribunal Superior do Trabalho		
Endereço Eletrônico (e-mail) leilio.bentes@tst.jus.br			Telefone fixo (61) 3043.4232		Telefone celular
Nome do Coordenador Projeto Alexandre William Majdalani					CPF 483.197.161-87
Unidade/Departamento CDS/SETIN			Cargo Coordenador de Desenvolvimento de Sistemas		
Endereço Eletrônico (e-mail) alexandre.majdalani@tst.jus.br			Telefone fixo (61) 3043-4427		Telefone celular (61) 98162-6340

#### 5. Descrição do Projeto

TÍTULO DO PROJETO	PERÍODO DE EXECUÇÃO
<b>Projeto SABIÁ</b> - Processamento de Linguagem Natural Aplicado ao Sistema Bem-Te-Vi para Análises de Processos Jurídicos do Tribunal Superior do Trabalho.	25 meses após a assinatura do instrumento.
<b>VALOR TOTAL</b> R\$ 2.262.981,01 (dois milhões, duzentos e sessenta e dois mil, novecentos e oitenta e um reais e um centavo)	

## 6. Introdução

As inovações promovidas pela Inteligência Artificial (IA) têm ocupado espaço privilegiado nos meios de comunicação. Conferências de negócios com vistas a aplicar a IA para aumentar a competitividade acontecem quase que diariamente pelo mundo. Tem sido cada vez mais comum, também, governos mundo afora publicando suas políticas para dominar a tecnologia. A IA tem assumido de forma inesperada o centro do discurso popular, muito em função dos resultados.

Os resultados alcançados pela IA tem potencial para mudar nossas vidas. A IA já figura como componente base de muitos dos sites e aplicações favoritas da sociedade. Só para nomear alguns: *Youtube, Facebook, Instagram, TikTok, Google Apps, Netflix, Amazon Alexa, Apple Siri, Waze, NuBank*. O seu uso foi importante para o desenvolvimento de vacinas e diagnóstico da COVID-19 ([Altan e Karasu, 2020], [Yang et al, 2021]).

Sensível à transformação promovida pela IA, o Poder Judiciário brasileiro tem envidado muitos esforços para adotar suas técnicas e ferramentas há pelo menos cinco anos. Alguns projetos de IA no judiciário ganharam muita repercussão recentemente. Dentre eles, destacam-se:

- **VICTOR**: projeto executado pelo AILAB, Dra. IA e STF com o objeto de dar suporte a seção judiciária do tribunal para classificar temas de repercussão. Além disso, o projeto desenvolveu solução para segmentação e classificação de peças jurídicas ([Braz et al, 2021], [Silva et al, 2019], [Araujo et al, 2020], [Hartmann e Bonat, 2019]).
- **Bem-Te-Vi**: O *Bem-Te-Vi* é um sistema para apoio à gestão e triagem do acervo dos Gabinetes de Ministro do TST. Utilizando tecnologias de big data, o software disponibiliza aos gabinetes informações sobre os processos de seu acervo de forma eficiente e acessível. Com o objetivo de complementar a análise do gabinete, são exibidas também informações produzidas por algoritmos de Inteligência Artificial que possibilitam, por exemplo, a denegação por transcendência, com base nos processos já julgados pelo magistrado. O sistema começou a funcionar nos gabinetes do Tribunal Superior do Trabalho (TST) em fevereiro de 2020. O maior ganho é na agilidade, que permite uma pesquisa rápida para gerenciamento dos processos. Ao possibilitar ao gabinete uma triagem de processos mais eficiente e precisa, também viabiliza a definição de estratégias para aumento da produção do gabinete. Para a sociedade, o sistema contribui para a redução do tempo de tramitação do processo e, conseqüentemente, para a celeridade processual.
- **Athos**: projeto executado pelo STJ com o objetivo de melhorar a efetividade de buscas processuais por meio de IA.
- **Mandamus**: projeto executado pelo AILAB, Dra. IA e TJRR com o objetivo de automatizar a geração de mandados a partir da leitura do despacho de magistrados; distribuição e monitoramento do cumprimento do mandado.
- **ALEI (Análise LEgal Inteligente)**: projeto executado pelo AILAB e TRF1 com o objetivo de auxiliar o analista judiciário no trabalho de instrução de minutas de voto, com referência à jurisprudência e precedentes, além da capacidade de sugestão de processos para voto em lote [Silva et al, 2019].



- **SINARA**: projeto executado pelo TRF3 com o objetivo de localizar entidades nomeadas (ER) relacionadas às classes: diploma, artigo, parágrafo, inciso, alínea, súmulas e precedentes.

Ciente da amplitude da mudança e das consequências proporcionadas pela IA, o CNJ tem se colocado como um orquestrador dessas iniciativas, fomentando a associação entre as entidades ativamente envolvidas nesse processo. Além disso, tem envidado ações para conferir uma regulação mínima, por meio de resoluções que expressem requisitos e preocupações mínimas a serem atendidas pelos projetos que apliquem IA. Por fim, destaca-se o fato de o CNJ estar trabalhando para proporcionar um ecossistema denominado Sinapses para que as diversas iniciativas possam ser compartilhadas entre os membros do Poder Judiciário [CNJ, 2020]. Dentre as iniciativas, destaca-se o CODEX, cujo objetivo é consolidar dados de todos os tribunais e disponibilizá-los tanto em texto, como o arquivo fonte para eventuais experimentos de IA, reduzindo assim o esforço de acesso ao dado, atividade que representa o principal gargalo de qualquer projeto.

O TST (protagonista com o projeto *Bem-Te-Vi*) e o AILAB (protagonista nos projetos *VICTOR*, *ALEI* e *Mandamus*), atentos aos desdobramentos contínuos da área de IA no âmbito judicial e, enxergando oportunidades concretas de pesquisas, juntam-se nesta proposta com o intuito de desenvolver novas soluções para evoluir ainda mais o sistema *Bem-Te-Vi* e, com isto, oferecer novas aplicações de IA para o TST e potencialmente para o Sistema Judiciário Brasileiro.

As próximas seções deste Plano de Trabalho detalham como a equipe do AILAB<sup>1</sup> pretende, com o apoio do grupo de pesquisas Dr IA<sup>2</sup> da Faculdade de Direito da UnB e em colaboração com equipe do TST, pesquisar, desenvolver e entregar importantes evoluções para o sistema *Bem-Te-Vi*.

## 7. Objeto e Escopo

O projeto tem como escopo pesquisar e desenvolver soluções para complementar o sistema *Bem-Te-Vi* no que se refere às funcionalidades de Agrupamento de Processos e Levantamento de Jurisprudência. Visando evidenciar o desempenho real destas soluções, é objetivo deste trabalho desenvolver um módulo de Registro de *Feedbacks* dos usuários. Por fim, este projeto também visa pesquisar soluções de *Long Life Machine Learning* (LLML) com o objetivo de desenvolver uma solução que traga um ciclo de vida longo aos modelos de IA implantados. Em síntese, este projeto visa o desenvolvimento de 2 (duas) soluções de IA, um sistema de registro de desempenho destas IAs e uma solução de LLML, conforme especificado abaixo:

1. Agrupar Processos: P&D de uma solução capaz de selecionar um subconjunto de processos similares a um processo de referência;
2. Identificar Jurisprudência: P&D de uma solução capaz de selecionar um subconjunto de decisões e acórdãos da base de jurisprudência que tenha correlação com um processo de referência;
3. Registrar *Feedbacks* de Usuários: Desenvolver um sistema de Registros de *Feedbacks* dos usuários quanto ao desempenho das soluções 1, e 2.

---

<sup>1</sup> <https://ailab.unb.br/>

<sup>2</sup> <http://dria.unb.br/>

4. Desenvolver um Subsistema de LLML: Desenvolver um subsistema que viabilize a semiautomação da atualização dos modelos de IA tendo os registros de *feedbacks* de usuários como referência.

## 8. Justificativa

Hoje, no governo federal brasileiro, é muito comum encontrar organizações que tentam melhorar seu desempenho, considerando, na maioria dos casos, soluções digitais e algumas aplicações tecnológicas. Não é difícil detectar atos normativos relacionados à governança digital. Somente para citar alguns (ordem cronológica), entre atos da Presidência da República e leis em consonância com o Congresso Nacional, foram selecionados:

- (1) *Eficiência organizacional*, Decreto nº 9.739 / 2019, que visa à administração pública federal ampliar e tratar o arranjo institucional do Sistema de Inovação do Governo Federal
- (2) *Rede Nacional de Governo Digital*, Decreto nº 9.584 / 2018, para promover a colaboração, o intercâmbio e a criação de iniciativas inovadoras relacionadas ao governo digital no setor público;
- (3) *Transformação digital*, Decreto nº 9.319 / 2018 que institui o Sistema Nacional para a Transformação Digital e estabelece a estrutura de governança para a implantação da Estratégia Brasileira para a Transformação Digital.
- (4) *Responsividade*, Decreto nº 9.203 / 2017 que dispõe sobre a política de governança da administração pública federal;
- (5) *Remoção de formalidades e soluções tecnológicas para simplificar o processo cidadão-cliente*, Lei Nacional 13.460 / 2017;
- (6) *Incentivo ao desenvolvimento científico, pesquisa, capacitação e inovação científica e tecnológica* ("Lei de Inovação"), Lei Nacional 13.243 / 2016;
- (7) *Política de governança digital*, Decreto nº 8.638 / 2016, que institui a Política de Governança Digital no âmbito da Administração Pública Federal.

A situação de morosidade e o congestionamento que caracterizam a estrutura judicial geram inevitável aumento de seus custos. Os avanços ocorridos em Tecnologia da Informação (TI) não têm sido capazes de superar os gargalos existentes e o constante congestionamento processual.

Uma via para ajudar o sistema de justiça, especialmente em relação aos processos em tramitação no TST, é a utilização de inovação tecnológica por meio de Aprendizado de Máquina (AM) no auxílio da realização de tarefas repetitivas.

Historicamente, Aprendizado de Máquina, subárea da Inteligência Artificial que evoluiu do estudo de reconhecimento de padrões e da teoria da aprendizagem computacional em inteligência artificial – tem se apresentado como alternativa metodológica para o enfrentamento de problemas de agrupamento (*clustering*), classificação, predição e inferência em grandes volumes de dados.

Em termos de inovação tecnológica, as entregas, resultado desta pesquisa e desenvolvimento, trarão melhorias importantes ao fluxo de análise processual nos Gabinetes de Ministros do Tribunal Superior do Trabalho (TST), otimizando o trabalho com o agrupamento de processos semelhantes e favorecendo a uniformização das decisões do TST.

Em termos de novos conhecimentos, colaboradores do TST, professores e acadêmicos dos cursos de Engenharia de Software e Direito da UnB terão a

oportunidade de se atualizarem com métodos e tecnologias inovadores hoje desenvolvidos no cenário internacional ao longo do desenvolvimento deste projeto interdisciplinar.

Em termos de produção científica na UnB, esta proposta prevê a participação de sua equipe em eventos relevantes da área do projeto, bem como a publicação dos resultados da pesquisa em veículos de comunicação científica.

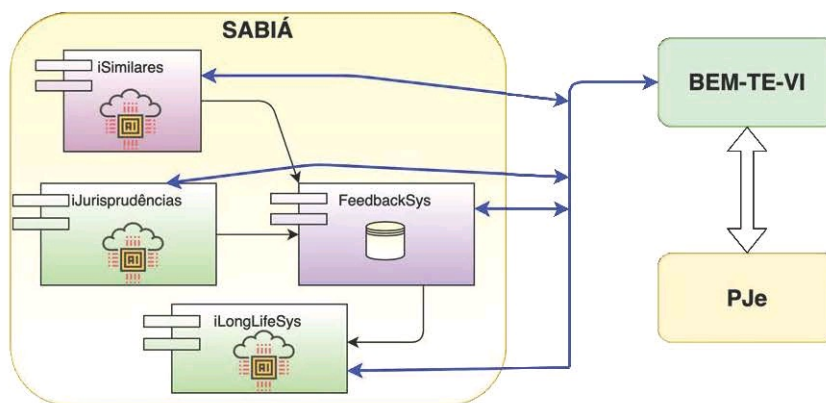
## 9. Método

O conhecimento teórico necessário para desenvolvimento da pesquisa será obtido por meio de revisão bibliográfica, recorrendo a livros, artigos, websites especializados, produções e experiências das equipes do IALAB, DR IA e do TST (TI, Unidades de Negócio e Desenvolvedores do *Bem-Te-Vi*).

As peças jurídicas de interesse e demais documentos processuais necessários ao desenvolvimento do projeto, bem como ao funcionamento dos entregáveis, serão fornecidos pelo TST.

Com o intuito de aproveitamento de experiências da equipe do AILAB e das características de integração com processamento paralelo, será adotada a linguagem de programação Python (<https://www.python.org/>) no desenvolvimento do sistema. Serão utilizadas bibliotecas especializadas em ML (Numpy, SciPy, Scikit-learn, Pandas, Matplotlib, entre outras) e frameworks para DNN (PyTorch, framework FastAI, HuggingFace e outros).

Os módulos de automação desenvolvidos serão disponibilizados no formato de microserviços a serem integrados ao Sistema *Bem-Te-Vi* do TST (Figura 1).



**Figura 1.** SABIÁ: Microserviços.

A pesquisa & desenvolvimento constará das seguintes metas:

### **M1. Compor a Base de Pesquisa e Desenvolvimento:**

Meta inicial que visa fornecer as condições para o desenvolvimento do projeto no que se refere à contratação da fundação de apoio, composição de equipes e disponibilização de infraestrutura para os trabalhos.



**M2. Desenvolver o módulo *iSimilares*:**

Pretende-se atender ao objetivo de agrupamento de processos com o módulo *iSimilares*. Este módulo deve ser capaz de relacionar processos similares a um determinado processo de referência, tendo metadados e/ou teores textuais como fontes de suas análises. Para tanto, será necessário validar as Hipóteses 1 e 2:

**Hipótese 1:** É possível desenvolver um sistema que priorize um subconjunto de processos jurídicos do acervo do TST a um determinado processo jurídico de interesse conforme um conceito de similaridade para fins de agrupamento de processos.

**Premissa:** O sistema *Bem-Te-Vi* fornecerá metadados e textos de peças jurídicas suficientes para se alcançar uma métrica de similaridade para fins de agrupamento de processos.

A questão se este módulo deverá atuar de forma geral ou estratificada (por gabinete, por exemplo) deverá ser definida junto com pessoal do TST.

**M3. Desenvolver o módulo *iJurisprudência*:**

Pretende-se atender ao objetivo de levantamento de jurisprudência com o módulo *iJurisprudência*. O conceito de jurisprudência aqui está relacionado a processos julgados do acervo do TST. Assim, este módulo deve ser capaz de correlacionar processos julgados a um determinado processo de referência, tendo metadados e/ou teores textuais como fontes de suas análises. Para tanto, será necessário validar a *Hipótese 3*:

**Hipótese 3:** É possível desenvolver um sistema que correlacione um subconjunto de processos julgados do acervo do TST a um determinado processo jurídico de interesse conforme um conceito de correlação para fins de levantamento de jurisprudência.

**Premissa:** O sistema *Bem-Te-Vi* fornecerá metadados e textos de peças jurídicas suficientes para se alcançar uma métrica de similaridade para fins de levantamento de jurisprudência.

A questão se este módulo deverá atuar de forma geral ou estratificada (por gabinete, por exemplo) deverá ser definida junto com pessoal do TST.

**M4. Desenvolver o módulo *FeedbackSys*:**

O módulo *FeedbackSys* visa atender ao objetivo de registrar feedbacks (avaliações) dos usuários sobre os desempenhos dos módulos com modelos de IA embarcados (*iSimilares*, *iJurisprudências*). Assim, este módulo deve oferecer meios para que o sistema *Bem-Te-Vi* envie registros de avaliações dos usuários. Estes registros deverão ser persistidos de forma adequada para o módulo *iLongLifeSys*.

**M5. Desenvolver módulo *iLongLifeSys*:**

O módulo *iLongLifeSys* visa atender ao objetivo de desenvolver um subsistema de LLML. Este subsistema deve viabilizar a semiautomação da atualização dos modelos de IA tendo os registros de feedbacks de usuários como referência.

#### **M6. Publicar Resultados das Pesquisas:**

Os resultados alcançados deverão ser submetidos para publicação em eventos e/ou revistas científicas, nacionais e/ou internacionais. O coordenador e pesquisadores envolvidos deverão liderar as ações desta meta de forma a oportunizar o trabalho colaborativo e coautorias com alunos e colaboradores do TST que se interessarem. Destaca-se ainda a importância da participação da equipe executora em eventos científicos da área, a fim de que o projeto possa ser retroalimentado com técnicas e conhecimentos novos relativos ao que outros grupos de pesquisas e empresas estão desenvolvendo de inovador na área de IA e PLN.

#### **M7. Transferir Tecnologia e Conhecimentos para TST:**

As transferências de tecnologias e de conhecimentos tem por objetivo viabilizar que o TST se aproprie dos resultados do projeto e consiga dar prosseguimento ao mesmo depois de sua entrega. Uma vez que o TST terá uma equipe técnica colaborando nas fases de P&D da proposta, espera-se que a apropriação das técnicas e das tecnologias desenvolvidas sejam paulatinamente internalizadas ao longo do desenvolvimento do projeto. Atividades específicas de transferências para usuários finais e outros colaboradores do TST também estão previstas – esta iniciativa visa a promoção do uso efetivo do projeto pelo público interno do TST.

Os artefatos de softwares (códigos-fonte, documentações), artefatos de dados (banco de dados, conjunto de dados para modelagem de IA), modelos de IA, bem como, documentações técnicas geradas deverão ser repassadas ao TST.

Esta pesquisa será desenvolvida por equipe interdisciplinar da UnB e com apoio de comissão técnico-gestora do TST. A UnB poderá contratar uma de suas fundações de apoio para a gestão das atividades administrativas e financeiras.

Preve-se que as metas M2, M3, M4, M5 e M6 devem direcionar as pesquisas em IA para as seguintes temáticas voltadas a Processamento de Linguagem Natural (PLN):

1. Classificação
2. Clustering
3. Long Life Machine Learning (LLML)

O trabalho [Vaswani et al, 2017] abriu perspectivas em processamento de linguagem natural (PNL) que, até então, eram inimagináveis. Os modelos decorrentes do método Transformer<sup>3</sup> proporcionaram resultados que revolucionaram várias tarefas, especialmente no que diz respeito a: tradução computadorizada, conversação, geração de conteúdo e classificação ([Devlin et al, 2019], [Radford et al, 2019], [Yang et al, 2019], [Liu et al, 2019], [Raffel et al, 2020]).

Alguns fatos são dignos de nota, quando nos referimos a modelos baseados em *Transformer*. Primeiramente, impressiona a elevada demanda de recursos computacionais para construir e operá-los. Isso é consequência da quantidade de parâmetros contemplados pela arquitetura das redes neurais, combinada ao expressivo

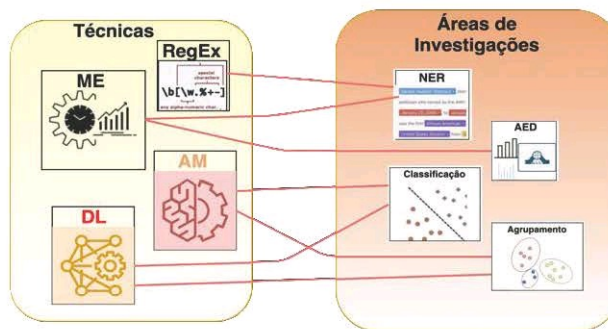
<sup>3</sup> O link <https://app.inferkit.com/demo> apresenta uma aplicação que faz uso de modelo baseado em Transformer, no caso, o GPT-2.

volume de dados usado para treiná-los. Ou seja, para se conseguir obter o modelo treinado, são necessárias dezenas *Graphic Processing Units* (GPUs) por vários dias. A NVIDIA, por exemplo, conseguiu treinar o BERT-Large em 47 minutos, usando 1.472 V100 GPUs [Liu et al, 2020]. A Tabela 1 exemplifica alguns desses modelos.

**Tabela 1.** Modelos de Linguagem.

Modelo	Parâmetros un. de milhão de amostras	Volume de Dados
BERT-Large	355	SQuAD 1.1 <sup>4</sup> 100.000 amostras
GPT-2	1.500	40 GB
GPT-3	175.000	570 GB

Outro fato relevante é que os modelos mais modernos de PNL citados anteriormente são treinados com dados em inglês. O que torna os seus modelos pré-treinados aptos a sofrerem adaptações para tarefas em sua língua nativa. Apesar de termos promovido ajuste fino em modelo originalmente em inglês para o português, observamos que ainda temos um caminho para que este modelo possa ser adaptado para uma área do domínio específico, como o direito, além de outras limitações associadas ao conjunto limitado de treinamento, para que fosse viável realizar o processo com os recursos que o AILAB conta no presente momento. Prevemos investigações relativas a Transformers [Vaswani et al, 2017b] e Transferência de Aprendizado em NLP [Raffel et al, 2020].



**Figura 2.** Técnicas a serem empregadas nas pesquisas.

Em termos de técnicas, prevemos aplicações de abordagens mais usuais como Expressões Regulares – RegEx na identificação de entidades nomeadas (*Named Entity Recognition* – NER) e Métodos Estatísticos (ME) principalmente para análises exploratórias dos dados (AED). Haverá investigações de potenciais usos de Métodos de Aprendizado de Máquina – AM, Aprendizado Profundo (*Deep Learning* - DL) e

<sup>4</sup> The Stanford Question Answering Dataset



*Transformers* para NER e na fase de modelagens em IA para realização de classificações e agrupamentos (Figura 2).

## 10. Resultados Esperados para TST

Espera-se que a consecução das metas e das investigações descritas atendam às expectativas do TST enumerada a seguir:

1. **Otimização dos recursos humanos envolvidos na condução das atividades administrativas e judiciais do TST:** como detalhado, as pesquisas e entrega de resultado por meio de microserviços que serão integrados ao sistema Bem-Te-Vi afetarão positivamente os colaboradores do TST que realizam atividades de análises processuais jurídicas.
2. **Aumento da capacidade de processamento do volume de demandas, o que permitirá tramitação mais veloz dos processos no TST:** a mesma automação citada anteriormente, operando em conjunto com o Bem-Te-Vi e PJe, elevará a capacidade de resposta desta corte aos usuários de seu sistema de julgamentos.
3. **Diminuição da Taxa de Congestionamento de processos no Judiciário:** com a elevação da produtividade decorrente da automação de tarefas repetitivas associadas a análises documentais e movimentação judicial, espera-se uma redução importante do congestionamento de processos trabalhistas.

## 11. Resultados Científicos e Tecnológicos Esperados:

Esperamos contribuir cientificamente com a sociedade interessada na temática de IA no judiciário com a divulgação científica dos resultados encontrados ao longo da execução do projeto. Para tanto, está prevista a submissão dos resultados a congressos e revistas de relevância para a área de IA e PLN.

O projeto Bem-Te-Vi do TST é um dos projetos pioneiros e de sucesso em aplicação de IA para o judiciário brasileiro. Os resultados do projeto SABIÁ contribuirão diretamente na incorporação de métodos e técnicas inovadoras e atuais ao sistema Bem-Te-Vi.

É esperado que soluções para agrupamento de processos similares e de levantamento de jurisprudência trabalhista sejam entregues contendo as técnicas e métodos mais atuais e promissores da área de Inteligência Artificial.

Os módulos de LLML e de Feedback de usuários se configuram em duas outras importantes tecnologias que deverão ser citadas como pioneiras em promoverem perenidade a sistemas de IA no judiciário brasileiro – um problema importante da área de IA cujas soluções estão sendo buscadas por importantes grupos de pesquisa em IA no Brasil e em outros países de expressiva produção tecnológica e científica.

O projeto em pauta traz uma proposta ousada do ponto de vista de investigações de assuntos de fronteira do conhecimento em IA e robusta quanto à sua proposta de entregar o resultado destas pesquisas avançadas na forma de soluções por softwares de uso direto e imediato para o sistema Bem-Te-Vi e TST.

## 12. Resultados Esperados para UnB

1. Qualificação de professores e discentes em área avançada de pesquisa em PLN.
2. Qualificação de professores e discentes em análise processual.
3. Qualificação de professores e discentes em processo de desenvolvimento de software ágil e DevOps.
4. Publicação científica dos resultados alcançados.
5. Melhoramento de recursos tecnológicos e de pessoal para projetos em IA.
6. Suporte ao desenvolvimento de trabalhos de conclusão de curso, dissertações de mestrado e teses de doutorado.

### 13. Cronograma de Execução

A execução do projeto ficará a cargo das equipes na Universidade de Brasília – UnB e de equipe com pessoal a ser designada pelo TST conforme detalhamento da tabela abaixo.

#### 13.1 Alteração no Cronograma de Execução

O Plano de Trabalho apresentado e aprovado constava do ordenamento de execução explícito no item 9 e detalhado no item 13. Entretanto, em reunião realizada no dia 24/02/22 foi proposto pela equipe do Tribunal Superior Trabalho um reordenamento das entregas, provocando a alteração das datas de execução entre as Meta 2, Meta 3 com impacto na Meta 4.

Tal solicitação se justifica pelo maior alcance que a implementação do **iJurisprudência** terá dentro do tribunal, provocando uso em cadeia dos módulos que, posteriormente, serão disponibilizados.

Dito isto, segue o cronograma de execução com o remanejamento de datas que preveem o desenvolvimento dos módulos **iJurisprudência**, **iSimilares** e demais reajustes.

Esta proposta resultou em Apostilamento aprovado em 24/03/23.

#### 13.2 Justificativa de alteração no Cronograma de Execução

O projeto surgiu com o objetivo de pesquisar e desenvolver soluções para complementar o sistema Bem-Te-Vi no que se refere às funcionalidades de Agrupamento de Processos e Levantamento de Jurisprudência. Visando evidenciar o desempenho real destas soluções, foi proposto o desenvolvimento de um módulo de Registro de Feedbacks dos usuários e a identificação de soluções de Long Life Machine Learning (LLML) que tragam um ciclo de vida longo aos modelos de IA implantados.

Para tanto, os estudos iniciaram-se com o ênfase no **iSimilares** objetivando a identificação de peças que poderiam ser utilizadas, bem como, na identificação de técnicas de IA possíveis de serem aplicadas no contexto. Isto feito, uma proposta de solução foi apresentada, tomando os estudos mais específicos e aprofundados em Long Life Machine Learning e métodos de feedback.

O resultado desse estudo foi o desenvolvimento das atividades 5, 6, 7 e 8 do Módulo **iSimilares** que, até então, estava em andamento conforme especificado abaixo:

- **Atividade 5:** revisão técnica da arquitetura e modelos de agrupamento de processos já implementados pelo TST.
- **Atividade 6:** revisão técnica e da literatura (RTL): Métodos e técnicas aplicáveis ao agrupamento de processos do TST.
- **Atividade 7:** formalizar o conceito de similaridade para fins de agrupamento.
- **Atividade 8:** definir o subconjunto de peças relevantes para agrupamento.

Até aqui, cumpria-se o cronograma de atividades previsto no Plano de trabalho do projeto de Pesquisa e Desenvolvimento. Neste, o módulo iSimilares foi previsto para ser a primeira solução a ser pesquisada pela equipe da UnB. Entretanto, em reunião realizada no dia 24/02/2022 foi proposto pela equipe do Tribunal Superior Trabalho um reordenamento das entregas o que ocasionou a inversão dos módulos de iSimilares com o iJurisprudência. Tal proposta foi aceita pela equipe da UnB e formalizada por meio do 1º Aditivo à TED 01/2021 TST/UNB.

Em reunião negocial no TST no dia 09/02/2023 com o feedback do TST evidenciou-se a necessidade de repensar o alvo definido ao longo das etapas de pesquisa executadas até então, que dava atenção à busca de soluções que considerassem o processo como um todo. Apesar desse direcionamento estar coerente com o plano de trabalho, com os instrumentos de análise de questões negociais gerados pela equipe da UnB e apreciados pela corte, houve entendimento geral de que seguir nesta direção levaria a resultados pouco úteis para o TST, visto que os representantes dos gabinetes expuseram com muita ênfase a sua necessidade de avaliação no âmbito dos assuntos discutidos em cada recurso, e não na perspectiva global.

Entendeu-se conjuntamente a necessidade de suspender as atividades técnicas de engenharia de IA para priorizar uma imersão nas experiências individuais dos gabinetes, sob a perspectiva de reconhecer lacunas no Bem-Te-Vi para alavancar a atividade de operação de triagem de recursos de cada gabinete. Os gabinetes considerados para a imersão foram:

- GMCB em 16/02/2023
- GMDS e GMKA em 02/03/2023;
- GMDS e GMALR em 06/03/2023;
- GMCMB em 09/03/2023

Ao longo de 3 semanas, a equipe de pesquisa do projeto realizou visitas aos gabinetes, conforme cronograma descrito acima. Tais visitas induziram análises mais amplas envolvendo tanto membros de frentes de negócio, quanto de engenharia de IA da UnB.

Como resultado desta imersão, o TST sugeriu reunir todos os esforços e retomar o módulo de iSimilares para construir uma proposta e somente após, voltar e repensar o módulo de iJurisprudência. A UnB, por sua vez, acatou e entendeu ser uma estratégia coerente já que: (1) o iSimilares possui funcionalidades que poderão ser aproveitadas em relação ao iJurisprudência e (2) o iJurisprudência precisa de um complemento e acréscimo de um mecanismo de controle de similaridade em seu resultado final.

O processo de construção e reconstrução de caminhos pode acontecer no decorrer de qualquer projeto de pesquisa e desenvolvimento, considerando que um projeto de pesquisa tem como objetivo a experimentação de métodos, exaurindo possibilidades, a fim de confirmar ou não as hipóteses propostas inicialmente. Isto posto, justifica-se a alteração das datas das atividades do plano de trabalho para refletir as mudanças propostas.

**Novo cronograma com atividades/indicadores físico-financeiro**

Meta	#	Atividade	Início	Fim	Respon sável	Entregáveis	MESES																							
							1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20				
							12/2021	1/2022	2/2022	3/2022	4/2022	5/2022	6/2022	7/2022	8/2022	9/2022	10/2022	11/2022	12/2022	1/2023	2/2023	3/2023	4/2023	5/2023	6/2023	7/2023				
	1	Contratação da Fundação de Apoio	1	1	UnB	Contratos	X																							
	2	Seleção e Contratação de pessoal	1	1	UnB	Contratos	X																							
M1	3	Aquisição de equipamentos	1	10	UnB	Notas Fiscais	X	X	X	X	X	X	X	X	X															
	4	Formar equipe técnica-gestora que acompanhará o projeto por parte do TST	1	2	TST	Ato de nomeação do TST	X	X																						
	5	Revisão técnica da arquitetura e modelos de agrupamento de processos já implementados pelo TST.	2	3	UnB/TST	Relatório técnico		X	X																					
	6	Revisão técnica e da literatura (RTL): Métodos e técnicas aplicáveis ao agrupamento de processos do TST.	3	5	UnB/TST	Relatório técnico			X	X	X																			
	7	Formalizar o conceito de similaridade para fins de agrupamento.	9	11	UnB/TST	Relatório técnico								X	X	X														
	8	Definir o subconjunto de peças relevantes para agrupamento.	10	12	UnB/TST	Relatório técnico									X	X	X													
	9	Construção de base de dados para modelagem de IA	16	20	UnB/TST	Base de dados no AILAB																X	X	X	X	X				
M2	10	Análise Exploratória de Dados (AED): Selecionar descritores a serem	16	23	UnB/TST	Relatório técnico																X	X	X	X	X				



[illegible]



[illegible]

Universidade de Brasília – UnB.



[illegible]

## 14. Plano de Aplicação (em Reais (R\$))

RECEITAS	
<b>A - Total de Receitas</b>	<b>2.262.981,01</b>
DESPESAS	
Especificação	Valor Total (R\$)
Diárias	24.315,00
Passagens e despesas com locomoção	25.127,51
Auxílio Financeiro a Pesquisador	1.525.000,00
Outros Serviços de Terceiros - Pessoa Jurídica	462.351,29
Equipamento e Material Permanente	217.000,00
<b>B - Total de Despesas</b>	<b>2.262.981,01</b>

## 15. Detalhamento das Despesas (em Reais (R\$))

Diária					
ITE M	Beneficiário	Cidade	Quant .	Valor unit.	Valor total
1	Diária Internacional	A Definir	15	1.365,00	20.475,00
2	Diária Nacional	A Definir	12	320,00	3.840,00
Subtotal					24.315,00

Passagem e despesa com locomoção					
ITE M	Beneficiário	Trecho	Quant .	Valor unit.	Valor total
1	Passagem Internacional	A Definir	3	6.375,84	19.127,51
2	Passagem Nacional	A Definir	4	1.500,00	6.000,00
Subtotal					25.127,51

Bolsa de Pesquisa						
ITE M	Beneficiário	Modalidade	Quant .	Período	Valor unit.	Valor total
1	Nilton Correia da Silva - (Área de Computação)	Pesquisador Sênior	1	25	9.000,00	225.000,00
2	Fabrizio Ataides Braz - (Área de Computação)	Pesquisador Sênior	1	25	9.000,00	225.000,00
3	Debora Bonat (Área de Direito)	Pesquisador Sênior	1	25	9.000,00	225.000,00
4	Luciana Nishi (Área de Computação)	Pesquisa, Desenvolvimento e Inovação	1	25	5.000,00	125.000,00

5	Aline Dayany de Lemos (Área de Computação)	Pesquisa, Desenvolvimento e Inovação	1	25	5.000,00	125.000,00
6	Jonathan Alis Salgado (Área de Computação)	Pesquisa, Desenvolvimento e Inovação	1	25	5.000,00	125.000,00
7	Pesquisador (Área de Computação)	Pesquisa, Desenvolvimento e Inovação	4	25	1.500,00	150.000,00
8	Bolsa de Pesquisa, Desenvolvimento e Inovação (Área de Computação)	Pesquisa, Desenvolvimento e Inovação	9	25	1.000,00	225.000,00
9	Bolsa de Pesquisa, Desenvolvimento e Inovação (Área de Direito)	Pesquisa, Desenvolvimento e Inovação	2	25	2.000,00	50.000,00
<b>Subtotal</b>						<b>1.525.000,00</b>

Outros Serviços de Terceiros - Pessoa Jurídica				
ITEM	Descrição	Quant.	Valor unit.	Valor total
1	Taxa de inscrição em evento científico internacional	3	3.000,00	9.000,00
2	Taxa de inscrição em evento científico nacional	4	500,00	2.000,00
3	Despesas Acessórias de Importação	1	42.000,00	42.000,00
4	Despesas Operacionais e Administrativas	1	174.557,04	174.557,04
5	Ressarcimento - FUB (resolução CAD nº 045/2014)	1	234.794,25	234.794,25
<b>Subtotal</b>				<b>462.351,29</b>

Equipamento e Material Permanente				
ITEM	Descrição	Quant.	Valor unit.	Valor total
1	Nobreak	1	7.000,00	7.000,00
2	Servidor GPU (IMPORTADO)	1	210.000,00	210.000,00
<b>Subtotal</b>				<b>217.000,00</b>

Imposto Sobre Serviços de Qualquer Natureza				
ITEM	Descrição	Quant.	Valor unit.	Valor total
1	Imposto Sobre Serviços (ISS)*	1	9.187,21	9.187,21
<b>Subtotal</b>				<b>9.187,21</b>
Encargo incorrido sobre a execução do projeto, com alíquota de 5%, conforme LC nº 116/2003, sobre o valor total do projeto, exceto ressarcimento da Instituição Federal de Ensino Superior (IFE).				

<b>B - Total de Despesas</b>	<b>2.262.981,01</b>
------------------------------	---------------------

Valores de bolsas conforme Resolução do Conselho de Administração da Universidade de Brasília Nº 0003/2018.

## 16. Cronograma de Desembolso



Mês 1: R\$ 1.424.323,69.

Mês 12: R\$ 838.657,32

## 17. Referências Bibliográficas

[Altan e Karasu, 2020] Aytaç Altan and Seçkin Karasu. Recognition of covid-19 disease from x-ray images by hybrid model consisting of 2d curvelet transform, chaotic salp swarm algorithm and deep learning technique. *Chaos, Solitons Fractals*, 140:110071, 2020.

[Araujo et al, 2020] Pedro Henrique Luz de Araujo, Teófilo Emídio de Campos, Fabricio Ataides Braz, and Nilton Correia da Silva. VICTOR: a dataset for Brazilian legal documents classification. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 1449– 1458, Marseille, France, May 2020. European Language Resources Association.

[Braz et al, 2021] Fabricio A. Braz, Nilton C. da Silva and Jonathan Silva. Leveraging effectiveness and efficiency in page stream deep segmentation. *International Scientific Journal Engineering Applications of Artificial Intelligence*, tbd:tbd, 2021.

[CNJ, 2020] Conselho Nacional de Justiça. Resolução nº332 de 21/08/2020. Online: <https://atos.cnj.jus.br/atos/detalhar/3429>, 21/08/2020.

[Devlin et al, 2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[Hartmann e Bonat, 2019] Fabiano Hartmann and Debora Bonat. Machine learning and the general repercussion on brazilian supreme court: Applying the victor robot to legal texts. In *MIREL@JURIX*, 2019.

[Liu et al, 2019] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

[Liu et al, 2020] Weijie Liu, Peng Zhou, Zhiruo Wang, Zhe Zhao, Haotang Deng, and Qi Ju. FastBERT: a self-distilling BERT with adaptive inference time. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6035–6044, Online, July 2020. Association for Computational Linguistics.

[Radford et al, 2019] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners, 2019.

[Raffel et al, 2020] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of

transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020.

[Silva et al, 2019] Nilton C. da Silva, Fabricio A. Braz, and José Roberto P. Ferretti. Aplicação da inteligência artificial no sistema jurídico brasileiro: Victor e Alei. *Revista de Administración Pública del GLAP*, 3:45–55, 2019.

[Vaswani et al, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[Vaswani et al, 2017b] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30 (NIPS)*, pages 5998–6008. Curran Associates, Inc., 2017.

[Yang et al, 2019] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alelle-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[Yang et al, 2021] Zikun Yang, Paul Bogdan, and Shahin Nazarian. An in silico deep learning approach to multi-epitope vaccine design: a sars-cov-2 case study. *Scientific Reports*, 11:3238, 02 2021.

MINISTRO LELIO BENTES CÔRREA  
PRESIDENTE  
TRIBUNAL SUPERIOR DO TRABALHO



Documento assinado eletronicamente por **LELIO BENTES CORRÊA, PRESIDENTE**, em 25/10/2023, às 11:53, conforme horário oficial de Brasília, com fundamento no inciso I do art. 4º da Lei nº 14.063, de 23 de setembro de 2020.



A autenticidade do documento pode ser conferida no site [https://sei.tst.jus.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.tst.jus.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0) informando o código verificador **0457462** e o código CRC **EC41BDDE**.